 Conduct all tests at $\alpha = 0.05$ significance level

All Questions are based on the following 2 regression models, where SIMPLE REGRESSION refers to the case where $p=1$, and $X$ is of full column rank (no linear dependencies among predictor variables).

Model 1: $Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + \epsilon_i$, $i = 1, \ldots, n$ $\epsilon_i \sim NID(0, \sigma^2)$

Model 2: $Y = X\beta + \epsilon$ $X = n \times p'$ $\beta = p \times 1$ $\epsilon \sim N(0, \sigma^2 I)$

Given: $\frac{d(a'x)}{dx} = a$ $\frac{d(x'Ax)}{dx} = 2Ax$ (A symmetric) $E(Y'AY) = tr(AV_v) + \mu_v' A \mu_v$

Cochran's Theorem: Suppose $Y$ is distributed as follows with nonsingular matrix $V$:

$Y \sim N(\mu, \sigma^2 V)$ $r(V) = n$ then if $AV$ is idempotent:

1. $Y' \left( \frac{1}{\sigma^2} A \right) Y$ is distributed non-central $\chi^2$ with: (a) $df = r(A)$ and (b) Noncentrality parameter: $\Omega = \frac{1}{2\sigma^2} \mu'A \mu$

2. $Y'AY$, $Y'BY$ independent if $AVB = 0$

---

**Critical Values for t and F-distributions**

F-distributions indexed by numerator df across top of table

<table>
<thead>
<tr>
<th>df</th>
<th>t(.05)</th>
<th>t(.025)</th>
<th>t(.01)</th>
<th>t(.005)</th>
<th>t(.001)</th>
<th>t(.0005)</th>
<th>f(.05,1)</th>
<th>f(.05,2)</th>
<th>f(.05,3)</th>
<th>f(.05,4)</th>
<th>f(.05,5)</th>
<th>f(.05,6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2.920</td>
<td>4.303</td>
<td>2.915</td>
<td>2.353</td>
<td>2.179</td>
<td>2.101</td>
<td>17.81</td>
<td>23.70</td>
<td>23.21</td>
<td>21.32</td>
<td>19.68</td>
<td>18.55</td>
</tr>
<tr>
<td>3</td>
<td>2.353</td>
<td>3.182</td>
<td>2.998</td>
<td>2.353</td>
<td>2.179</td>
<td>2.101</td>
<td>15.83</td>
<td>21.32</td>
<td>20.52</td>
<td>18.55</td>
<td>16.75</td>
<td>15.51</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

---

**Critical Values for t and F-distributions**

F-distributions indexed by numerator df across top of table

<table>
<thead>
<tr>
<th>df</th>
<th>t(.05)</th>
<th>t(.025)</th>
<th>t(.01)</th>
<th>t(.005)</th>
<th>t(.001)</th>
<th>t(.0005)</th>
<th>f(.05,1)</th>
<th>f(.05,2)</th>
<th>f(.05,3)</th>
<th>f(.05,4)</th>
<th>f(.05,5)</th>
<th>f(.05,6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2.920</td>
<td>4.303</td>
<td>2.915</td>
<td>2.353</td>
<td>2.179</td>
<td>2.101</td>
<td>17.81</td>
<td>23.70</td>
<td>23.21</td>
<td>21.32</td>
<td>19.68</td>
<td>18.55</td>
</tr>
<tr>
<td>3</td>
<td>2.353</td>
<td>3.182</td>
<td>2.998</td>
<td>2.353</td>
<td>2.179</td>
<td>2.101</td>
<td>15.83</td>
<td>21.32</td>
<td>20.52</td>
<td>18.55</td>
<td>16.75</td>
<td>15.51</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

---

**Critical Values for t and F-distributions**

F-distributions indexed by numerator df across top of table

<table>
<thead>
<tr>
<th>df</th>
<th>t(.05)</th>
<th>t(.025)</th>
<th>t(.01)</th>
<th>t(.005)</th>
<th>t(.001)</th>
<th>t(.0005)</th>
<th>f(.05,1)</th>
<th>f(.05,2)</th>
<th>f(.05,3)</th>
<th>f(.05,4)</th>
<th>f(.05,5)</th>
<th>f(.05,6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2.920</td>
<td>4.303</td>
<td>2.915</td>
<td>2.353</td>
<td>2.179</td>
<td>2.101</td>
<td>17.81</td>
<td>23.70</td>
<td>23.21</td>
<td>21.32</td>
<td>19.68</td>
<td>18.55</td>
</tr>
<tr>
<td>3</td>
<td>2.353</td>
<td>3.182</td>
<td>2.998</td>
<td>2.353</td>
<td>2.179</td>
<td>2.101</td>
<td>15.83</td>
<td>21.32</td>
<td>20.52</td>
<td>18.55</td>
<td>16.75</td>
<td>15.51</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Q.1. A study considered noise level of the Teheran-Karaj express train (Y, in dB) in terms of distance to the center of the track (X₁, in meters) and speed of the train (X₂, in km/h), with n = 50. Consider the following models (in matrix form).

\[ \text{Model 0: } E\{Y\} = \beta_0 \ \ \ \ \text{Model 01: } E\{Y\} = \beta_0 + \beta_1 X_1 \ \ \ \ \text{Model 02: } E\{Y\} = \beta_0 + \beta_2 X_2 \]

\[ \text{Model 012: } E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \quad Y'Y = 348654 \]

p.1.a. The sum of the speeds of the 50 observations is \( \sum_{i=1}^{50} Y_i = 4174.2 \). For model 0, obtain:

\[ X_0'X_0, \quad X_0'Y, \quad (X_0'X_0)^{-1}, \quad \hat{\beta}_0, \quad Y'P_0Y \]

\[ X_0'X_0 = 50 \quad X_0'Y = 4174.2 \quad (X_0'X_0)^{-1} = \frac{1}{50} \quad \hat{\beta}_0 = \frac{83.484}{50} \quad Y'P_0Y = 348478.9 \]

p.1.b. For Models 01, 02, and 012, you obtain the following

\[ X_1'Y, \ \ \ \ \hat{\beta}_1, \ \ \ \ \hat{\beta}_2, \ \ \ \ \hat{\beta}_{12}, \ \ \ \ \hat{\beta}_{01}, \ \ \ \ \hat{\beta}_{02}, \ \ \ \ \hat{\beta}_{12} \]

\[ X_1'Y, \ \ \ \ \hat{\beta}_1, \ \ \ \ \hat{\beta}_2, \ \ \ \ \hat{\beta}_{12}, \ \ \ \ \hat{\beta}_{01}, \ \ \ \ \hat{\beta}_{02}, \ \ \ \ \hat{\beta}_{12} \]

\[ Y'P_0Y = 348478.9 \quad Y'P_{01}Y = 348499.6 \quad Y'P_{02}Y = 348627.6 \quad MSE_{012} = 0.562 \]

p.1.c. Obtain the Sequential and Partial sums of squares for X₁ and X₂, and their corresponding F-statistics.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Sequential SS</th>
<th>Sequential F</th>
<th>Partial SS</th>
<th>Partial F</th>
</tr>
</thead>
<tbody>
<tr>
<td>X₁</td>
<td>228.5</td>
<td>228.5</td>
<td>128.0</td>
<td>228.5</td>
</tr>
<tr>
<td>X₂</td>
<td>36.1</td>
<td>36.1</td>
<td>20.3</td>
<td>36.1</td>
</tr>
</tbody>
</table>

\[ R(\beta_1 | \beta_0) = 607.3 - 478.9 = 128.4 \]

\[ R(\beta_1 | \beta_0, \beta_2) = 627.6 - 499.6 = 128.0 \]

\[ R(\beta_2 | \beta_0, \beta_1) = 627.6 - 607.3 = 20.3 \]

\[ R(\beta_2 | \beta_0, A) = 20.3 \]
Q.2. Consider the general linear test $H_0: \mathbf{K}'\beta = 0$ where $\mathbf{K}'$ has $q \leq p'$ linearly independent rows.

p.2.a. Derive the mean vector and variance-covariance matrix of $\mathbf{K}'\hat{\beta}$.

$$
\begin{align*}
\mathbf{E}\{\mathbf{K}'\hat{\beta}\} &= \mathbf{K}'\mathbf{E}\{\hat{\beta}\} = \mathbf{K}'\beta \\
\mathbf{V}\{\mathbf{K}'\hat{\beta}\} &= \mathbf{K}'\mathbf{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K} = \mathbf{\sigma}^2(\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K})
\end{align*}
$$

p.2.b. Show that $\mathbf{Q} = \begin{pmatrix} \mathbf{K}' \mathbf{\beta} \end{pmatrix} \begin{pmatrix} \mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{K} \end{pmatrix}^{-1} \mathbf{K}'\mathbf{\beta}$ and $\text{SSE}$ are independent. Hint: Write $\mathbf{Q} = \mathbf{Y}'\mathbf{A}\mathbf{Y}$.

$$
\begin{align*}
\text{SSE} &= \mathbf{Y}'(\mathbf{I}-\mathbf{P})\mathbf{Y} \\
\mathbf{V}\{\mathbf{Y}\} &= \mathbf{\sigma}^2 \mathbf{I} = ? \mathbf{V} = \mathbf{I}
\end{align*}
$$

$$
\begin{align*}
\mathbf{Q} &= \mathbf{\hat{\beta}}'\mathbf{K}[\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}]^{-1}\mathbf{K}'\mathbf{\hat{\beta}} = \mathbf{Y}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}[\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}]^{-1}\mathbf{K}'\mathbf{Y} \\
(\mathbf{I}-\mathbf{P})(\mathbf{I}-\mathbf{P}) &= \mathbf{I} - 2\mathbf{P} + \mathbf{P}\mathbf{P} = \mathbf{I} - 2\mathbf{P} + \mathbf{P} = \mathbf{I} - \mathbf{P} \\
\mathbf{A}\mathbf{U}\mathbf{A}' &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}[\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}]^{-1}\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}[\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}]^{-1}\mathbf{K}'\mathbf{X}' (\mathbf{I}-\mathbf{P})
\end{align*}
$$

$$
\begin{align*}
\mathbf{I} &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}[\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}]^{-1}\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{A}\mathbf{V} \\
\mathbf{A}\mathbf{V}\mathbf{B} &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}[\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}]^{-1}\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' (\mathbf{I}-\mathbf{P})
\end{align*}
$$

$$
\begin{align*}
\mathbf{X}'(\mathbf{I}-\mathbf{P}) &= \mathbf{X}' - \mathbf{X}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{X}' - \mathbf{X}' = 0 \\
\text{Independent}
\end{align*}
$$
Q.3. A study was conducted, relating an abrasivity index measure (Y) to \( p = 4 \) predictors: UCS (X_1), BTS (X_2), and two brittleness indices: B_1 (X_3) and B_2 (X_4) in a sample of igneous rocks. The model fit is given below along with computations.

\[
Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \epsilon_i \quad \epsilon_i \sim NID(0, \sigma^2)
\]

<table>
<thead>
<tr>
<th>X'X</th>
<th>40</th>
<th>102.81</th>
<th>51.975</th>
<th>62.766</th>
<th>93.597</th>
<th>109.60</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>102.81</td>
<td>121.9227</td>
<td>80.71724</td>
<td>93.99107</td>
<td>136.4427</td>
<td>152.32</td>
</tr>
<tr>
<td></td>
<td>51.975</td>
<td>121.9227</td>
<td>80.71724</td>
<td>93.99107</td>
<td>136.4427</td>
<td>152.32</td>
</tr>
<tr>
<td></td>
<td>62.766</td>
<td>121.9227</td>
<td>80.71724</td>
<td>93.99107</td>
<td>136.4427</td>
<td>152.32</td>
</tr>
<tr>
<td></td>
<td>93.597</td>
<td>121.9227</td>
<td>80.71724</td>
<td>93.99107</td>
<td>136.4427</td>
<td>152.32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>X'Y</th>
<th>331.28</th>
<th>300.30</th>
<th>322.05</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>INV(X'X)</th>
<th>1.509</th>
<th>-0.071</th>
<th>1.554</th>
<th>-0.634</th>
<th>-0.994</th>
<th>5.671</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.071</td>
<td>0.023</td>
<td>0.045</td>
<td>-0.017</td>
<td>-0.008</td>
<td>-0.255</td>
</tr>
<tr>
<td></td>
<td>1.554</td>
<td>0.045</td>
<td>3.028</td>
<td>-1.150</td>
<td>-1.624</td>
<td>4.972</td>
</tr>
<tr>
<td></td>
<td>-0.634</td>
<td>-0.017</td>
<td>-1.150</td>
<td>0.482</td>
<td>0.605</td>
<td>-1.305</td>
</tr>
<tr>
<td></td>
<td>-0.994</td>
<td>-0.008</td>
<td>-1.624</td>
<td>0.605</td>
<td>0.930</td>
<td>-2.859</td>
</tr>
</tbody>
</table>

| Beta-hat Y/INV(X'X)Y | 331.28 | 300.30 | 322.05 |

p.3.a. Compute SSE, df_e and MSE

\[
\text{SSE} = 331.28 - 322.05 = 9.23
\]

\[
df_e = 35
\]

\[
\text{MSE} = \frac{9.23}{35} = 0.264
\]

p.3.b. Compute SSR, df_r and MSR

\[
\text{SSR} = 322.05 - 300.30 = 21.75
\]

\[
df_r = p - 1 = 4
\]

\[
\text{MSR} = \frac{21.75}{4} = 5.438
\]

p.3.c. Test \( H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0 \) (abrasivity index is not associated with any of the predictors)

\[
F_{0.05, 4, 35} = \frac{\text{MSR}}{\text{MSE}} = \frac{5.438}{0.264} = 20.60
\]

Test Statistic \( F = 20.60 \)

Rejection Region \( F \geq 2.648 \)

\( P > 0.05 \)

p.3.c. Test \( H_0: \beta_3 = \beta_4 \) (The coefficients for the two brittleness indices are equal)

\[
K' = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}
\]

\[
k = \begin{bmatrix} 0.360 & -0.009 & 0.474 & -0.123 & -0.325 \end{bmatrix}
\]

\[
Q = (1.554)^2 \begin{bmatrix} 0.202 \end{bmatrix} = 11.96
\]

\[
F = \frac{Q}{1} \frac{1}{0.264} = \frac{11.96}{0.264} = 45.3
\]

Test Statistic \( F = 45.3 \)

Rejection Region \( F \geq 4.128 \)

\( P > 0.05 \)
Q.4. An experiment was conducted relating springiness in berries \((Y_y, \text{ in mm})\) to sugar equivalent \((X, \text{ in g/L})\). There were \(c = 4\) distinct sugar equivalent groups, with \(n_j = 5\) berries per group. The lack-of-fit test for a linear relation is:

\[H_0 : E\{Y_{ij}\} = \mu_j = \beta_0 + \beta_1 X_j \quad i = 1, ..., 5; \quad j = 1, ..., 4 \quad H_A : E\{Y_{ij}\} = \mu_j = \beta_0 + \beta_1 X_j \]

The ANOVA table is provided:

<table>
<thead>
<tr>
<th>term</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>(F)</th>
<th>significance (F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>245.74</td>
<td>245.74</td>
<td>57.40</td>
<td>0.0000</td>
</tr>
<tr>
<td>Residual</td>
<td>18</td>
<td>77.06</td>
<td>4.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>322.80</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>i,j</th>
<th>(X_j)</th>
<th>(n_j)</th>
<th>(\hat{y}_{ij})</th>
<th>(y_{ij})</th>
<th>(s^2_{ij})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>176.5</td>
<td>5</td>
<td>21.89</td>
<td>0.38</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>209.3</td>
<td>5</td>
<td>18.04</td>
<td>18.77</td>
<td>2.90</td>
</tr>
<tr>
<td>3</td>
<td>225</td>
<td>5</td>
<td>16.20</td>
<td>15.41</td>
<td>11.30</td>
</tr>
<tr>
<td>4</td>
<td>259.5</td>
<td>5</td>
<td>12.14</td>
<td>12.32</td>
<td>3.18</td>
</tr>
</tbody>
</table>

Coefficient Standard Error \(t\) Stat \(P\)-value

| Intercept  | 42.6025 | 3.4019 | 12.5233 | 0.0000 |
| susCont    | -0.1174 | 0.0155 | -7.5763 | 0.0000 |

Note: \(s^2_j = \frac{\sum_{i=1}^{n_j}(Y_{ij} - \bar{Y}_j)^2}{n_j - 1}\) \(n_j > 1, 0\) otherwise

\(n = 20, c = 4\)

p.4.a. Give the fitted value for the linear regression for the 4th group \((X_4 = 259.5)\).

\[
\hat{Y}_{44} = 42.6025 - 0.1174(259.5) = 12.14
\]

p.4.b. Compute the Pure Error Sum of Squares, degrees of freedom and Mean Square.

\[
SSPE = \sum_{j=1}^{c} (n_j-1) S_j^2 = 4 \left[0.38 + 2.90 + 11.50 + 3.18\right] = 4(17.76) = 71.04
\]

\[
df_{PE} = n - c = 20 - 4 = 16
\]

\[
MSPE = \frac{71.04}{16} = 4.44
\]

p.4.c. Compute the Lack-of-Fit Sum of Squares, degrees of freedom and Mean Square.

\[
SS_{LF} = \sum_{j=1}^{c} n_j \left(\frac{\sum_{i=1}^{n_j}(Y_{ij} - \hat{Y}_{ij})^2}{n_j - 1}\right) = 5\left(\frac{(-0.12)^2 + 0.73^2 + (-0.79)^2 + 0.18^2}{n_j - 1}\right)
\]

\[
= 5(1.2038) = 6.02
\]

\[
df_{LF} = c - 2 = 4 - 2 = 2
\]

\[
MSLF = \frac{6.02}{2} = 3.01
\]

p.4.d. Give the Test Statistic, Rejection Region, and P-value relative to .05 for the Lack-of-Fit test.

\[
F_{MSLF/MSPE} = 3.01 = 0.68
\]

Test Statistic: \(F_{LF} = 0.68\)  Rejection Region: \(F_{LF} \geq 3.634\)  P-value > 0.05
Q.5. Regression models were fit, relating various crime rates for U.S. states to a set of 25 predictors. The researchers fit the full model with all 25 predictors (say \(X_1, ..., X_{25}\)) and then the best 4 predictor model (say \(X_1, ..., X_4\)). For the outcome Total Crime, the authors report the following coefficients of multiple determination.

\[
R^2(X_1, ..., X_{25}) = .913 \quad R^2(X_1, ..., X_4) = .774
\]

Compute \(R^2(X_5, ..., X_{25} | X_1, ..., X_4)\)

\[
R^2(X_5, ..., X_{25} | X_1, ..., X_4) = \frac{R^2(X_1, ..., X_{25}) - R^2(X_1, ..., X_4)}{1 - R^2(X_1, ..., X_4)} = \frac{.913 - .774}{1 - .774} = \frac{.139}{.226} = .615
\]

Q.6. A regression model is fit, relating energy consumption (Y) to total area (X) for a sample of \(n = 19\) luxury hotels in Hainan Province, China. The Analysis of Variance for the simple linear regression model is given below.

<table>
<thead>
<tr>
<th>ANOVA</th>
<th>(df)</th>
<th>(SS)</th>
<th>(MS)</th>
<th>(F)</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regressio</td>
<td>1</td>
<td>25521.76</td>
<td>25521.76</td>
<td>57.75</td>
<td>0.0000</td>
</tr>
<tr>
<td>Residual</td>
<td>17</td>
<td>7512.94</td>
<td>441.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>18</td>
<td>33034.70</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

p.6a. A plot of the residuals versus area is given below. It demonstrates which possible violations of assumptions (circle all that apply).

Non-normal Errors \[\text{Unequal Variance}\] Serial Correlation of Errors Non-linear Relation between Y and X

p.6b. A second regression model is fit, relating the squared residuals (Y) to area (X). Conduct the Breusch-Pagan test to test whether the equal variance assumption is reasonable. The sums of squares are given below.

\[
\begin{align*}
\chi^2_{BP} &= \frac{SSR^2X}{(SSe)^2} = \frac{12393791}{(7512.94)^2} \\
&= \frac{619689.5}{1563553.3} = 3.963
\end{align*}
\]

Test Statistic: \(\chi^2_{BP} = 3.963\)

Rejection Region: \(\chi^2_{0.05,1} = 3.84\)

P-value > or \(\leq\) 0.05
Q.7. A regression model was fit based on a sample of n=117 Black Holes. The response was Bolometric Luminosity (Y), with predictors: Black Hole Mass (X_1) and Black Hole Type (X_2 = 1 if Radio Quiet Quasar (RQQ), 0 if Radio Loud Quasar (RLQ)), and a cross-product term to allow for a possible interaction between Mass and Type. Note that there were 20 RQQ and 97 RLQ Black Holes.

\[
\hat{Y}_1 = 39.356 + 0.791X_{11} + 2.047X_{12} - 0.257X_{11}X_{12} \quad \text{SSE}_1 = 30.964
\]

\[
\hat{Y}_2 = 39.453 + 0.780X_{11} \quad \text{SSE}_2 = 31.187
\]

p.7.a. Based on Model 1, give the fitted equations relating Bolometric Luminosity to Mass, separately by Quasar Type.

RQQ: \( \hat{Y} = 41.403 + 0.534X_{11} \)  

RLQ: \( \hat{Y} = 39.356 + 0.791X_{11} \)

p.7.b. Test whether the true relationship between Bolometric Luminosity and Mass is the same for RQQs and RLQs.

\[
H_0: \beta_2 = \beta_3 = 0 \quad H_a: \beta_2 \text{ and/or } \beta_3 \neq 0
\]

T.S. \( F_{663} = \frac{31.187 - 30.964}{30.964} \times \frac{113}{113} = 0.407 \)

Test Statistic: \( F = 0.407 \)  

Rejection Region: \( F \geq 3.077 \)  

P-value > or < .05

p.7.c. When the regressions are fit separately, the fitted equations are the same as you should have in part p.7.a. The residual variances (MSE's) for the models are: RQQ: s_1^2 = 0.088  

RLQ: s_2^2 = 0.309. Use Bartlett's test to test whether the true variances are equal.

\[
B = \frac{1}{C} \left[ \nu \ln(MSE) - \sum \nu_i \ln(s_i^2) \right] = 1.01 \quad \text{N} = 113
\]

\[
C = 1 + \frac{1}{3(4-1)} \left[ \sum \nu_i - \nu \right] = 1.01 \quad \text{N}_1 = 20 - 2 = 18
\]

\[
\nu = 97 - 2 = 95
\]

\[
113 \ln(0.088) - (18 \ln(0.088) + 95 \ln(0.309)) = (-146.29) - \left[ -43.74 + (-111.57) \right] = -146.29 + 155.31 = 9.02
\]

\[
\beta = \frac{9.02}{1.01} = 8.931
\]

Test Statistic: \( \chi^2 = 8.931 \)  

Rejection Region: \( \chi^2 \geq 3.841 \)  

P-value > or < .05
Q.8. A random sample of \( n = 15 \) Bollywood movies was obtained, and a simple linear regression was fit relating the log Revenues (\( Y \)) to log Budget (\( X \)) in international markets. Movie 15 was Sultan with log Revenues \( Y_{15} = 5.705 \) and log Budget \( X_{15} = 4.500 \). Regression models were fit with and without Sultan, and results are given below.

<table>
<thead>
<tr>
<th>XX</th>
<th>XY</th>
<th>X'X_(15)</th>
<th>X'Y_(15)</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.000</td>
<td>55.474</td>
<td>53.796</td>
<td>14.000</td>
</tr>
<tr>
<td>55.474</td>
<td>214.736</td>
<td>213.087</td>
<td>50.974</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INV(XX)</th>
<th>Beta-hat</th>
<th>INV((X'X)_(15))</th>
<th>Beta-hat_(15)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.495</td>
<td>-0.386</td>
<td>-1.871</td>
<td>1.563</td>
</tr>
<tr>
<td>-0.386</td>
<td>0.104</td>
<td>1.476</td>
<td>-0.410</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>YY</th>
<th>YPY</th>
<th>YY_15</th>
<th>YPY_15</th>
</tr>
</thead>
<tbody>
<tr>
<td>218.04</td>
<td>213.79</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

p.8.a. Obtain SSE and MSE and the estimated error standard deviation for each model.

\[
SSE_1 = 218.04 - 213.79 = 4.25 \\
SSE_2 = 185.50 - 182.26 = 3.24
\]

\[
\text{df}_e = 15 - 2 - 13 \quad \text{MSE}_1 = 0.327 \quad \sigma = 0.572
\]

\[
\text{df}_e = 14 - 2 - 12 \quad \text{MSE}_2 = 0.270 \quad \sigma = 0.520
\]

p.8.b. Obtain the fitted value for Sultan for each model and its leverage value based on the full data set.

\[
\hat{Y}_1 = -1.871 + 1.476 \left( 4.5 \right) = 4.771 \quad \left[ \begin{array}{c} 1 \\ 4.5 \end{array} \right] \left[ \begin{array}{c} 1.476 \\ -0.386 \end{array} \right] = \left[ \begin{array}{c} 4.771 \\ 4.624 \end{array} \right]
\]

\[
\text{Leverage} = 0.127 \quad n_{15,15} = 4.5
\]

p.8.c. Compute DFFITS\(_{15}\) for Sultan. Note that DFFITS makes use of \( \sigma_{(15)} \) to estimate \( \sigma \).

\[
\text{DFFITS}_{15} = \frac{4.771 - 4.624}{0.520 \sqrt{0.127}} \quad u(\hat{Y}) = \sigma^2 P
\]

\[
\sigma_{(15)} = 0.185 \quad u(\sigma_{(15)}) = \sigma^2 \quad DFFITS_{15} = \frac{0.795}{0.185} = 4.26 \quad DFFITS_{15} = 0.795
\]